

EVOLUCIÓN DEL CONCEPTO DE INTELIGENCIA ARTIFICIAL (IA) Y SUS CONSECUENCIAS JURÍDICAS

EVOLUTION OF THE CONCEPT OF ARTIFICIAL INTELLIGENCE (AI) AND ITS LEGAL CONSEQUENCES

ENRIQUE VÁZQUEZ PITA

Doctorando del programa de Derecho y Ciencias Sociales
de la UNED

evazquez56@alumno.uned.es

Sumario: *I. Introducción. II. Evolución de la definición técnica de la IA. II.A. El seminario de Dartmouth y sus críticos. II.B. La visión cognitivista y conexionista. II.C. Aprendizaje y autonomía. II.D. Tecnología de la predicción. II.E. Críticas a la definición de IA III. Sistemas de IA III.A Tipos de sistemas de IA: IA General (fuerte), Débil (estrecha) y Singleton (extrema). IV. Las lagunas en la AI Act. V. Conclusiones. VI. Bibliografía.*

Resumen: La primera definición oficial de Inteligencia Artificial (IA) se remonta al Seminario de verano de Dartmouth en 1956, que enunció esta tecnología como “los sistemas que demuestran un comportamiento inteligente analizando su entorno y actuando, con cierto grado de autonomía, para lograr objetivos específicos”. Sin embargo, su formulación ha evolucionado a medida que sus prestaciones técnicas han mejorado y dejaron obsoleta la definición previa. Actualmente, la AI Act o Ley de Inteligencia Artificial define el sistema de IA como “un sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía y que puede mostrar adaptabilidad después del despliegue y que, para objetivos explícitos o implícitos, infiere, a partir de la información que recibe, cómo

generar resultados como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales”. Sin embargo, algunos autores escépticos cuestionan el propio nombre de IA porque lo consideran engañoso o incorrecto o lo rechazan porque ni es inteligente ni es artificial. La indefinición de una tecnología de riesgo sistémico como la IA plantea cuestiones jurídicas de calado porque las innovaciones, como la IA generativa y otras, generan sucesivas lagunas legales y le restan eficacia a la ley. Ello, probablemente, exigirá una revisión continua y puesta al día de la definición jurídica de la AI Act.

Palabras clave: Inteligencia Artificial (IA), legislación, AI Act, definición, laguna legal.

Abstract: The first official definition of Artificial Intelligence (AI) dates back to the Dartmouth Summer Seminar in 1956, which defined this technology as “systems that demonstrate intelligent behavior by analyzing their environment and acting, with some degree of autonomy, to achieve specific goals”. However, its formulation has evolved as its technical capabilities have improved, rendering the previous definition obsolete. Currently, the AI Act defines an AI system as “a machine-based system designed to operate with varying levels of autonomy and capable of exhibiting adaptability after deployment, and that, for explicit or implicit objectives, infers from input how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.” However, some skeptical authors question the very name AI because they consider it misleading or incorrect, or reject it because it is neither intelligent nor artificial. The lack of definition of a systemically risky technology like AI raises significant legal issues because innovations, such as generative AI and others, create successive legal loopholes and render the law less effective. This will likely require ongoing review and updating of the legal definition of the AI Act.

Key words: Artificial Intelligence (AI), legislation, AI Act, definition, loophole.

I. INTRODUCCIÓN

Una de las dificultades que plantea la IA es que no hay acuerdo sobre su definición, ni en el ámbito internacional ni en el mundo académico. La UE ha optado en su Libro Blanco por una definición que

describe lo que hace la IA¹, pero esa táctica implica que la definición sea provisional porque cada vez se le encuentran más usos y habilidades a esta tecnología emergente que presenta un crecimiento exponencial. Posteriormente, en su Propuesta de Reglamento (conocida como Ley de la IA o AI Act), la UE admite la idea de complementar la definición de esta tecnología con una lista de nuevas aplicaciones que surjan y el artículo 3 del Proyecto Final de la AI Act de mayo de 2024 la identifica con una herramienta basada en la adaptabilidad que genera predicciones, recomendaciones o decisiones que influyen en un entorno físico o virtual (esto último pensando en el metaverso). La indefinición de la IA tiene consecuencias desde el punto de vista jurídico, al no concretar el objeto de estudio. Como advierte MERABET, S.² se están empleando numerosas denominaciones para designar las tecnologías inteligentes (IA, robot, machine, ordenador, algoritmo, big data, sistemas expertos, agente artificial y otras) pero no todas designan una misma realidad. Alerta de que, en el Estado de Derecho positivo, la ausencia de designaciones únicas de las tecnologías inteligentes y las carencias de las definiciones usadas contribuyen a aumentar la incertidumbre. Y hay que recordar que el Derecho Positivo, al contrario que el Natural, se basa en preceptos y estos han de ser precisos, que no necesariamente correctos.

Como prueba de la evolución constante de la definición de la IA, la asociación Elements of AI recalca que hace 50 años se consideraba que los métodos automáticos de búsqueda y planificación pertenecían al ámbito de la IA pero, en el 2018, esos métodos se enseñan a cualquier estudiante de informática. Según su argumentación, ahora, la IA es definida como un método para gestionar información incierta, pero, en unos años, esa disciplina se transferirá al ámbito de la estadística o la probabilidad.

Desde el punto de vista de la cibernética, la IA es una disciplina encuadrada dentro de las ciencias cognitivas formadas por las neurociencias, la lingüística computacional, la antropología cognitiva, la psicología cognitiva, y la filosofía de la cognición.³

¹ La introducción del Libro Blanco especifica que «en otras palabras, la inteligencia artificial es una combinación de tecnologías que agrupa datos, algoritmos y capacidad informática». Y añade que es «una de las partes más importantes de la economía de los datos».

² MERABET, S., *Vers un droit de l'intelligence artificielle*. Nouvelle Bibliothèque de Thèses, v. 197, Éditions Dalloz, París, 2020, pág. 49.

³ MENECEUR, Y., *L'intelligence artificielle en procès*, Editorial Bryuland, Bruselas, 2020, pág. 46.

MENECEUR distingue tres etapas en el desarrollo de la IA. Entre 1940 y 1960 se inicia la disciplina de la IA al abrigo de la cibernética. Entre 1980 y 1990, priman los sistemas expertos y la investigación mediante la fuerza bruta. Después del 2010, surge un nuevo esplendor de la IA a partir del tratamiento de los datos masivos (el Big Data) y de una novedosa potencia de cálculo (GPU).⁴

En 2020, el ALLAI (organismo independiente holandés fundado por los tres miembros holandeses del Grupo de Expertos de Alto Nivel en IA de la UE) dividió en cuatro las definiciones de la IA: 1) una tecnología, 2) un próximo paso en la digitalización, por el cual la opinión es que todo es IA, 3) un campo de la ciencia destinado a modelar la inteligencia como medio para comprender la inteligencia y 4) una herramienta o entidad «mágica» omnisciente, todopoderosa, que nos sucede sin que tengamos ningún poder para controlarla⁵.

En los últimos años han surgido voces críticas que rechazan el término de inteligencia artificial y su narrativa, caso de autores como MOROZOV, E., LARSON, E.J., o LEUFER, D. y también a aquellas como ZUBOFF, S. o CRAWFORD, K., que han ampliado el concepto de la IA hasta considerarlo un complejo ecosistema o un «constructo social» diseñado para extraer datos de los usuarios e influir en su conducta. Finalmente, surge el concepto de «agente inteligente».

II. EVOLUCIÓN DE LA DEFINICIÓN TÉCNICA DE LA IA

La Real Academia de la Lengua Española define inteligencia artificial como «la disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico»⁶.

La Enciclopedia Británica define IA como «la habilidad de una computadora digital o una computadora que controla a un robot para resolver problemas que son normalmente asociados con el más alto intelecto y capacidad de procesamiento de los humanos».⁷

La Wikipedia⁸ define la IA como «una rama de la informática que desarrolla máquinas y software con inteligencia similar a la huma-

⁴ MENECEUR (2020), págs. 21, 47-53.

⁵ ALLAI. «Primer análisis del Libro Blanco de la UE sobre IA». <https://allai.nl/first-analysis-of-the-eu-whitepaper-on-ai/>

⁶ Véase acepción de «inteligencia» en <https://dle.rae.es/inteligencia>

⁷ Citado por ERTEL, W., *Introduction to Artificial Intelligence*, Springer, Cham (Suiza), 2011, 2017, pág. 2.

⁸ Wikipedia.

na». Desde el punto de vista de la informática, como «la inteligencia expresada por máquinas, sus procesadores y sus programas de software, que serían los análogos al cuerpo, el cerebro y la mente, respectivamente, a diferencia de la inteligencia natural demostrada por humanos y ciertos animales con cerebros complejos». Y desde la perspectiva de la computación, la IA, considerada como una máquina «inteligente» ideal, es «un agente flexible que percibe su entorno y lleva a cabo acciones que maximicen sus posibilidades de éxito en algún objetivo o tarea».

II.A. El seminario de Dartmouth y sus críticos

El campo científico sobre la computación electrónica y las máquinas pensantes arranca con TURING, A. después de impartir su conferencia titulada «¿Puede pensar una máquina?»⁹ en 1947 y, más tarde, en 1950, tras publicar su artículo «Computing machinery and intelligence»¹⁰ donde presenta el «juego de simulación», conocido como test de Turing (test Captcha), donde una máquina intenta engañar a un jurado para que decidan si su interlocutor es un humano o no.¹¹ A mediados de la década de 1950, los expertos adoptaron la primera definición técnica de inteligencia artificial. McCARTHY, J., pionero de la computación, acuñó el término de Inteligencia Artificial en un seminario de verano en 1956 en Dartmouth. Definió la IA como «los sistemas que demuestran un comportamiento inteligente analizando su entorno y actuando, con cierto grado de autonomía, para lograr objetivos específicos».¹² Estableció que «la meta de la IA es desarrollar máquinas cuyo comportamiento haga pensar que ellas son inteligentes».¹³

⁹ TURING, A., ¿Puede pensar una máquina?, KRK Ediciones, Oviedo, 2012.

¹⁰ TURING, A.M., «Computing Machinery and Intelligence», *Mind*, Volume LIX, Issue 236, 1 de octubre de 1950, págs. 433–460.

¹¹ En el año 2014, por primera vez, el chatbot de Eugene Gootsman, logró vencer a treinta jueces que estaban participando en la prueba de que estaban chateando con un niño ucraniano de trece años. SANGUINETTI, P. replica que la superación de la prueba de Turing no impresionó a la comunidad de expertos en IA porque solo evidencia que dicho test es un criterio insuficiente (2023), págs. 57-58.

¹² Citado por CASTETS-RENARD, 2020, pág. 325. Hace referencia al artículo «Artificial Intelligence: A European Perspective», Joint Research Centre, EUR 29425 EN, 2018.

https://ai-watch.ec.europa.eu/publications/artificial-intelligence-european-perspective_en

¹³ Citado por ERTEL, W., *Introduction to Artificial Intelligence*. Springer: Cham (Suiza), 2011, 2017, pág.1. La cita en inglés es: «The goal of AI is to develop machines that behave as though they were intelligent».

Esta primera definición que considera la IA como «sistemas» ha tenido un enorme impacto en el siglo XXI dado que la AI Act o las Recomendaciones de la UNESCO para la gobernanza de la IA asumen que la inteligencia artificial son «sistemas» y perpetúan varias de las características enumeradas en los años 50 del siglo XX: autonomía, análisis del entorno y logro de objetivos. La definición de la AI Act, como se verá, es heredera de la visión de Dartmouth, la cual pretendía distanciarse del concepto de cibernética.

MCCARTHY, J. organizó en 1956 la Conferencia de Dartmouth (Dartmouth Summer Research Project on Artificial Intelligence)¹⁴, la cual definió la IA y plantó el germen de esta industria tecnológica. Estableció un proyecto para que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia pueda ser descrita con tanta precisión que pueda fabricarse una máquina para simularlo. Pretendían averiguar cómo fabricar máquinas que utilizasen el lenguaje, formasen abstracciones y conceptos, resolviesen las clases de problemas ahora reservados para los seres humanos y mejorasen por sí mismas. La propuesta, firmada por MCCARTHY, J., MINSKY, M., ROCHESTER, N. y SHANNON, C.,¹⁵ ya incluía la creación de computadoras, procesamiento del lenguaje natural, redes neuronales, teoría de la computación, abstracción y creatividad. Sostenían que las diferencias entre humanos y máquinas eran «ilusorias». La autora CASTETS-RENARD (2020) concluye que, dentro de este contexto de la conferencia de Dartmouth, «la inteligencia se debe entonces al hecho de que la máquina imita las funciones cognitivas asociadas con la mente humana o animal, que son las capacidades de aprendizaje y resolución de problemas»¹⁶.

El filósofo DREYFUS, H. criticó el argumento de dichos ingenieros porque «ni siquiera consideran la posibilidad de que el cerebro procese la información de una manera completamente diferente a la de una computadora».¹⁷ Por su parte, ULLMAN, E. ha deplorado este dualismo cartesiano de la IA, el hecho de creer que la mente es como una computadora, un «pecado original» que considera que «la IA se entiende estrictamente como inteligencia incorpórea, alejada de cual-

¹⁴ Al encuentro, según Stottler Henke Associates, asistieron Ray Solomonoff, Oliver Selfridge, Trenchard More, Arthur Samuel, Herbert A. Simon y Allen Newell.

¹⁵ Otros autores (CRAWFORD, 2021) incluyen en la misma reunión a HOOVER, G., LICKLIDER, J.C.R., NEWELL, A., SIMON, H. y WIENER, N.

¹⁶ CASTETS-RENARD, C., *Droit du marché unique numérique et intelligence artificielle. «Droit de l'Union européenne. Manuels»*. Editorial Bruylant, Bruselas, 2020, pág. 325.

¹⁷ Citado por CRAWFORD, K., *Atlas of AI*, Yale University Press, New Haven y Londres, 2021, pág. 6 DREYFUS es citado en *What Computers Can't Do*.

quier relación con la palabra material».¹⁸ Otro autor, WEIZENBAUM, J., afirmó que la idea de pensar en los humanos como meros sistemas de procesamiento de información es una noción demasiado simplista de inteligencia y que impulsó la «perversa gran fantasía» de que un científico de IA podría crear una máquina que aprende «como lo hace un niño».¹⁹

RICH, E. matizó la cuestión y consideró que «la IA es el estudio de cómo las computadoras hacen cosas en las que, por el momento, la gente es mejor».²⁰

Otra definición tradicional es la de «un método que busca formas susceptibles de dotar a los sistemas informáticos de capacidades intelectuales comparables a las de los humanos».²¹

MICHIE, D. redefinió la IA en 1978 como «un conocimiento que puede producir una confiabilidad y competencia de codificación que sobrepasa con creces el nivel más alto que el experto humano haya alcanzado sin ayuda alguna vez, tal vez incluso nunca».²²

II.B. La visión cognitivista y conexionista

Todas estas definiciones de mediados y finales del siglo XX se quedaron desfasadas a causa de lo que WEBB, A. denomina la «extraña paradoja»: cuando las técnicas inventadas pasan a ser de uso corriente, «se vuelven invisibles para nosotros». Concluye que «en su forma básica, la IA es un sistema que toma decisiones autónomas. Las tareas que lleva a cabo replican o imitan los actos de la inteligencia humana, tales como reconocer sonidos y objetos, resolver problemas, comprender el lenguaje y usar estrategias para alcanzar metas».²³

MENECEUR (2020) divide las posteriores definiciones de la IA en dos grupos: la visión simbólica o cognitivista, que considera la IA como sistemas expertos, y que se remonta a los años 1970-1990 y la

¹⁸ Citado por CRAWFORD, K., Yale University Press, Atlas of AI. New Haven y Londres, 2021, pág. 7. ULLMAN es citada en Life in Code, págs. 136-137.

¹⁹ Citado por CRAWFORD, K. (2021), pág. 5. Cita extraída de WEIZENBAUM, Computer Power and Human Reason, pág. 202-203.

²⁰ Citado por ERTEL, W. (2011, 2017), pág.2.

²¹ Citado por MERABET, La Recherche, enero de 1979, n. 96, v. 10, pág. 61.

²² Citado por CRAWFORD, K., Atlas of AI, New Haven y Londres, Yale University Press, 2021, pág. 7. La cita fue extraída de Quoted in Gill, Artificial Intelligence for Society, pág. 3.

²³ WEBB, A., Los nueve gigantes. Cómo las grandes tecnológicas amenazan el futuro de la humanidad, Ediciones Península, Barcelona, 2021, pág. 29.

visión conexionista²⁴, que supuso un cambio de paradigma pues considera la IA como un aprendizaje automático y estadístico basado en la retroalimentación y la predicción, y que arranca en el 2010 y hasta la fecha.²⁵

RUSSELL, S. y NORVIG, P. (2009) consideran que, ante todo, un sistema de IA es «racional»²⁶ como un complejo capaz de «operar de forma autónoma, percibir el entorno, persistir durante un período de tiempo prolongado, adaptarse al cambio y crear y perseguir (el mejor resultado esperado)»²⁷. Algunos autores (TZIMAS, Th.) los catalogan como funcionalistas. CRAWFORD matiza que ambos autores, concretamente, dijeron que «la inteligencia se ocupa principalmente de la acción racional. Idealmente, un agente inteligente toma la mejor acción posible en una situación».²⁸

BODEN, M. A., en 2016, señala que la IA «tiene como objeto que los ordenadores hagan la misma clase de cosas que puede hacer la mente». Aunque algunas, como razonar, se describen como inteligentes, otras como la visión, no, pero «todas entrañan competencias psicológicas que permiten a los seres humanos y a los animales alcanzar sus objetivos». Añade que la IA utiliza muchas técnicas diferentes para resolver una gran variedad de tareas. Tiene dos objetivos: uno es tecnológico (usar los ordenadores para hacer cosas útiles, a veces empleando métodos muy distintos a la mente) y el otro científico (usar conceptos y modelos de IA que ayuden a resolver cuestiones sobre los seres humanos y demás seres vivos)²⁹.

BODEN, M.A. aporta un concepto novedoso al considerar la IA como una «máquina virtual» o más bien, un sistema compuesto de varias máquinas virtuales. Aunque la IA necesita de máquinas físicas (un ordenador), «sería más acertado considerar que utiliza lo que los especialistas llaman máquinas virtuales». Define una máquina virtual como el «sistema de procesamiento de la información que el programador concibe cuando escribe un programa y el que tiene en mente la gente al usarlo». Suelen estar compuestas por patrones de actividad

²⁴ Entre los conexionistas o expertos en redes neuronales figuran HINTON, G., LeCUN, Y. Y BENGIO, Y.

²⁵ Ídem, MENECEUR, Y. (2020), págs. 20-21.

²⁶ RUSSELL, S. Y NORVIG, P., *Artificial Intelligence: A Modern Approach*, Prentice Hall, New Jersey, Tercera edición, 2021, Citado por FERNÁNDEZ CARBALLO-CALERO (2021), pág. 24.

²⁷ RUSSELL y NORVIG, (2010).

²⁸ Citado por CRAWFORD, K. (2021), pág. 7. La cita procede de RUSSEL, S. Y NORVIG, P. *Artificial Intelligence*, pág.30

²⁹ BODEN, M.A., *Inteligencia artificial*, Turner Publicaciones, Madrid, 2022, págs. 11-12

(o procesamiento de información) a varios niveles o capas. También considera máquinas virtuales a los lenguajes de programación.³⁰

KAPLAN, J. (2017) concluye que sobre las definiciones sobre IA no hay acuerdo y además «hay pocas razones para creer que la inteligencia maquina tenga demasiada relación con la inteligencia humana, al menos hasta ahora». Según indica, la mayoría de las definiciones confluyen en una única acepción: «Crear programas de ordenador o máquinas capaces de conductas que consideraríamos inteligentes si las efectuasen humanos», aunque cree que esta definición adolece de «grandes defectos», por ejemplo, al medir la inteligencia humana o introducir sesgos culturales y diferentes contextos.³¹

II.C. Aprendizaje y autonomía

En 2017, la Declaración de Montreal para un Desarrollo Responsable de la IA definió la IA como «una serie de técnicas que permiten que una máquina imite el aprendizaje humano, en concreto para aprender, hacer predicciones, tomar decisiones y percibir lo que la rodea. En el caso de un sistema informático, la inteligencia artificial se aplica a los datos digitales».³²

En 2018, la guía Elements of AI³³, cursos gratuitos creados por Reaktor y la Universidad de Helsinki, en Finlandia, admitió que no hay una definición exacta y el campo se va redefiniendo a medida que ciertos temas se clasifican como ajenos a la IA y surgen otros nuevos. Se fija en dos características clave de la IA: la autonomía y la adaptabilidad. Respecto a la autonomía, es la capacidad de ejecutar tareas en situaciones complejas sin la dirección constante del usuario. En cuanto a la adaptabilidad, es la capacidad de mejorar la ejecución de las tareas aprendiendo de la experiencia. Considera que la IA es un método, «una disciplina científica, un conjunto de conceptos, problemas y métodos para resolver los problemas».

³⁰ BODEN, M.A., Inteligencia artificial, Turner Publicaciones, Madrid, 2022, págs. 13-14

³¹ KAPLAN, J., Inteligencia Artificial. Lo que todo el mundo debe saber, Teell Editorial, 2021, págs. 1-3.

³² Glosario de la Declaración de Montreal para un Desarrollo Responsable de la IA (visto 16/12/2021) https://www.declarationmontreal-iaresponsable.com/_files/ugd/ebc3a3_e0eed6b6841a4ba3a711427f67c8aa8b.pdf

³³ Elements of IA. <http://course.elementsofai.com>

En 2019, la guía del Oxford Internet Institute (OII) de la Universidad de Oxford en colaboración con Google ³⁴definió la IA como «la programación de ordenadores que aprende y se adapta». Añade que la IA es «la denominación de cualquier sistema informático al que se enseña a reproducir comportamientos humanos inteligentes».

Ese mismo año, ROSSI, F. ³⁵, de IBM, definió la inteligencia como la habilidad de lograr metas en un amplio rango de ambientes o entornos. Y la inteligencia artificial como «inteligencia en un agente artificial».

BRIGHTON, H. y HOWARD, S. (2019) consideran que el Santo Grial de la IA es comprender al hombre como una máquina³⁶. Además, la IA también intenta alcanzar una teoría general de la acción inteligente en «agentes»: no simplemente para humanos o animales, sino para individuos en un sentido más amplio. Definen «agentes» como algo capaz de un comportamiento inteligente, como un robot (agente físico) o un programa de ordenador (agente virtual o de software). Los agentes virtuales ocasionalmente adquieren una ejemplificación física al descargarse en el cuerpo robótico. Un agente puede estar compuesto de subagentes.

II.D. Tecnología de la predicción

AGRAWAL, A. (2019)³⁷ aporta una definición técnica de la IA pero antes aclara que la ola actual de avances en IA no aporta inteligencia sino un componente crítico de la misma: la predicción. Por ello, define la IA como «una tecnología de la predicción».³⁸ Define la predicción como la capacidad para aportar la información que falta. La predicción toma la información que poseemos, a la que se suele hacer referencia como «datos» y la usa para generar información que no poseemos. Añade que la predicción, ahora con una caída continua de su coste, es una información básica en la toma de decisiones, que se puede combinar con la teoría de la lógica de la decisión. Advierte que esta tecnología implica una serie de compromisos y concesiones:

³⁴ Oxford Internet Institute (OII). <https://atozofai.withgoogle.com/intl/es/artificial-intelligence/>

³⁵ ROSSI, F., «AI Ethics for Enterprise AI». IBM Research AI, 2019. <https://www.youtube.com/watch?v=9YrQLSAyNs0>

³⁶ BRIGHTON, H. y SELINA, H., Inteligencia Artificial. Una guía ilustrada, Tecnos, Madrid, 2019, pág. 12.

³⁷ AGRAWAL, A., Gans, J. y Goldfarb, A. Máquinas predictivas. La economía simple de la Inteligencia Artificial, Editorial Reverte, Barcelona, 2019.

³⁸ AGRAWAL, A., (2019), pág. 3.

a más velocidad, menor precisión; a más autonomía, menor control; cuantos más datos, menor privacidad³⁹.

AGRAWAL, A. considera que los avances recientes en aprendizaje de máquinas se describen como avances de la inteligencia artificial porque aprenden y mejoran con el tiempo, producen predicciones más precisas que otros enfoques bajo ciertas condiciones y la mejora en la precisión de la predicción de estos sistemas les permite realizar tareas, tales como traducir o conducir, que antes estaban reservadas para los seres humanos⁴⁰.

AYLLÓN GARCÍA, J.D. indica que, desde un punto de vista científico, entendemos por inteligencia artificial a las técnicas de procesamiento de información caracterizadas por hacer cálculos sobre determinada información en un espacio dimensional virtual y construido mediante operaciones llevadas a cabo dentro del propio algoritmo para aprovechar diversas propiedades de espacios altamente dimensionales.⁴¹

BODEI, R. (2019) considera la IA como «una disciplina que estudia los fundamentos teóricos y que crea algoritmos destinados a la producción de un software capaz de imitar algunos tipos de inteligencia humana —aplicados a varios dispositivos mecatrónicos, algunos de los cuales, como los robots, provistos de ordenadores, motores, actuadores (piernas, brazos, «manos», ruedas), sensores y equipamiento auxiliar— con el objetivo de solucionar problemas concretos o de conseguir determinados fines».⁴²

El economista NIÑO-BECERRA, S. (2020) define la IA como «inteligencia exhibida por máquinas» y se aplica a una máquina que imita las funciones cognitivas que los humanos asocian con otras mentes humanas, como, por ejemplo, aprender y resolver problemas.⁴³

MERABET, S. (2020)⁴⁴ considera que las actuales definiciones de IA no son satisfactorias porque no tienen en cuenta criterios como que la «cosa» sea libre, tenga autonomía y voluntad y genere incertidumbre, lo que la asemeja en parte a la inteligencia humana. Avanza

³⁹ Idem AGRAWAL, A., pág. 5.

⁴⁰ Idem AGRAWAL, A., pág. 49.

⁴¹ AYLLÓN GARCÍA, J.A., «La inteligencia artificial como un medio para administrar Justicia», en Fodertis 8.0. Estudios sobre tecnologías disruptivas y justicia, Editorial Comares, Granada, 2020, pág. 5.

⁴² BODEI, R., Dominio y sometimiento. Esclavos, animales, máquinas, inteligencia artificial, Alianza Editorial, Madrid, 2022, págs. 250-251.

⁴³ NIÑO-BECERRA, S., Capitalismo (1679-2065), Editorial Ariel, Planeta, Barcelona, 2020, págs. 404-405.

⁴⁴ MERABET, S. (2020), pág. 120.

su definición de la IA como «aquella inteligencia que consiste en un sistema informático dotado de capacidades cognitivas que le permiten efectuar cosas de manera autónoma, que no son determinadas por la persona que la creó o que la usa».

A mayor abundamiento, MERABET detecta lagunas respecto a la identificación científica porque es dual. Por un lado, la IA es definida como «la búsqueda de medios susceptibles de dotar a los sistemas informáticos de capacidades intelectuales comparables a las humanas».⁴⁵ Para el autor, el algoritmo es ahora una pieza fundamental en los sistemas informáticos, es su ADN, y sin él no hay base suficiente para identificar a la IA. Sin embargo, esta definición técnica no debe reducir la IA a un mero sistema informático y propone que vaya asociada a una vertiente jurídica⁴⁶ para comprender la realidad de la IA, pues su noción es elusiva.

Por su parte, RUSSELL, S. (2020) cree que hay un problema en afinar con la definición básica de la IA. Cree que la clave consiste en entender que las máquinas inteligentes no realizan acciones para conseguir «sus» objetivos sino para lograr «nuestros» objetivos. Para este autor, lo más correcto es definir la IA como «máquinas que son beneficiosas en la medida en que se puede esperar que sus acciones logren nuestros objetivos». La cuestión es que «nuestros» objetivos se refiere a las preferencias e intereses de 8.000 millones de humanos, con toda su enorme variedad.⁴⁷

FERNÁNDEZ CARBALLO-CALERO, P., (2021)⁴⁸ coincide en que no existe una definición universal del término «inteligencia artificial». Explica que, generalmente, se considera «una disciplina de la informática cuyo objetivo es el desarrollo de máquinas y sistemas que puedan llevar a cabo tareas para las que se prevé la necesidad de intervención de inteligencia humana». Debido a que el aprendizaje automático y el aprendizaje profundo (Deep Learning) son dos esferas dentro de la IA, CARBALLO-CALERO matiza que, con el desarrollo de nuevas técnicas basadas en redes neuronales y de hardware, «solemos referirnos a la IA como sinónimo de aprendizaje automático profundo supervisado».

⁴⁵ MERABET, S., (2020), pág.53. op. cit., La Recherche, enero de 1979, n. 96, vol. 10, pág.61.

⁴⁶ Idem. MERABET, S., (2020), págs. 55-56.

⁴⁷ RUSSELL, S., Human Compatible. AI and Problem of Control, Penguin, Gran Bretaña, 2020, págs. 11-12.

⁴⁸ FERNÁNDEZ CARBALLO-CALERO, P.,, La propiedad intelectual de obras creadas por inteligencia artificial, Aranzadi-Thomson Reuters, Cizu Menor (Navarra), 2021, pág. 175.

Una de las más recientes definiciones procede de CRAWFORD, K. (2021), cofundadora del AI Now Institute de la New York University⁴⁹. La autora rehuye una definición técnica y ve la IA como un «constructo político y social» que refleja los intereses del poder. «La IA no es ni artificial ni inteligente. Más bien, la inteligencia artificial está incorporada y es material, hecha de recursos naturales, combustible, trabajo humano, infraestructuras, logística, historias y clasificaciones. Los sistemas de inteligencia artificial no son autónomos, racionales o capaces de discernir nada sin un entrenamiento extenso e intensivo en computación con grandes conjuntos de datos o reglas y recompensas predefinidas. De hecho, la IA, tal y como la conocemos, depende por completo de un conjunto mucho más amplio de estructuras políticas y sociales. Y debido al capital requerido para construir IA a escala y las formas de ver que optimiza, los sistemas de IA están diseñados en última instancia para servir a los intereses dominantes existentes. En este sentido, la inteligencia artificial es un registro de poder», dice la autora.⁵⁰

La definición de CRAWFORD de la IA como un «constructo social» ayudaría a explicar el surgimiento de diversos fenómenos que acompañan a los algoritmos y a los agentes inteligentes, tales como los sesgos o la discriminación o la vulneración de otros derechos universales como la dignidad o la intimidad, y que se abordarán en los próximos epígrafes. En la misma línea, WEBB, A. sostiene que el hecho de que los programadores sean hombres, blancos, con estudios en universidades de élite y de clase acomodada provoca inevitablemente un sesgo de la IA hacia la visión del mundo de esta «tribu», que tiene sus propias reglas y rituales.⁵¹ Es lo que PANIAGUA, E. define como «brogrammers» («brothers» y «programmers», combinación de hermanos y programadores). Se trataría de colegao o hermandad del colectivo de informáticos masculinos, que considera como una cultura machista que expulsa a las mujeres a través de micromachismos⁵².

En 2021, el programa de IA GPT-3, desarrollado por OpenAI de Amazon para completar textos automáticamente en el proceso de escritura, definió la Inteligencia Artificial como «un conjunto de sistemas de orden superior capaces de razonar, planificar y resolver problemas. Y una vez que una máquina o un proceso está definido como

⁴⁹ AI Now Institute at New York University: <https://ainowinstitute.org/>

⁵⁰ CRAWFORD, K., *Atlas of AI*, Yale University Press, New Haven y Londres, 2021, pág. 8.

⁵¹ WEBB, A. (2021), págs. 92-93.

⁵² PANIAGUA, E., *Error 404. ¿Preparados para un mundo sin Internet?*, Penguin Random House Grupo Editorial, Barcelona, 2021, pág. 169.

inteligente, la IA se manifiesta en varios procesos, procesos que se utilizarán [...]».⁵³ En sentido estricto, se trata de la primera definición de IA elaborada por un agente inteligente no humano.

En 2022, Open AI, en aquel momento propiedad de Elon Musk hasta que dejó la empresa, y cuando desarrollaba programas como GPT-3, DALL-E 2 o OpenAI Codex, definió la AGI (Inteligencia Artificial General) como «sistemas altamente autónomos que superan a los humanos en el más económico y evaluable trabajo».⁵⁴

II.E. Críticas a la definición de IA

CABALLERO, R. y MARTÍN, E (2022) se acogen a esta definición: «La IA es el campo de la informática que busca replicar tareas complejas que habitualmente han venido realizando humanos, incluyendo reconocimiento de imágenes, escritura o voz, o incluso toma de decisiones en tiempo real, como las que deben tomar los vehículos autónomos».⁵⁵

Sin embargo, CABALLERO y MARTIN hacen posteriormente esta reflexión: «¿Podemos hablar realmente de inteligencia artificial?». Recuerdan que ante todo se trata de «métodos que aprenden a partir de datos». Estos modelos son «superespecializados» y «cortos de miras», por lo que, según los autores, no se prevé que hagan grandes descubrimientos o inferencias interesantes.⁵⁶

Por su parte, el proyecto A Is For Another: A Dictionary of AI indica en el 2020 que «la IA es un conjunto de tecnologías que incluyen el aprendizaje automático (ML), la visión computerizada, el razonamiento, y procesamiento del lenguaje natural, entre otros. Existe en un espacio incómodo y único como tecnología, metáfora e imaginario sociotécnico».⁵⁷

⁵³ SANTANA VEGA, C., «Esta Inteligencia Artificial ESCRIBE MEJOR QUE TÚ... ¡Y puedes probarla! GPT-3», Canal Dot CSV en YouTube. https://youtu.be/C1eOiOkD_8A

⁵⁴ OpenAI's mission is to ensure that artificial general intelligence (AGI) —by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity. We will attempt to directly build safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome. <https://openai.com/charter/>

⁵⁵ CABALLERO, R. y MARTÍN, E., *Las bases de big data y de la inteligencia artificial*, Los libros de la Catarata, Madrid, 2022, pág. 97.

⁵⁶ Ídem, CABALLERO y MARTÍN, págs. 115-116.

⁵⁷ <https://www.donestech.net/noticia/another-dictionary-ai>

En noviembre de 2022, Open AI lanzó ChatGPT, un bot conversacional que contestaba preguntas con un saber enciclopédico. En solo cinco días, logró un millón de usuarios y a principios del 2023 superaba los cien millones de usuarios. Eso llevó a MOROZOV, E. a criticar las definiciones de inteligencia artificial al considerar que «ni es inteligente ni es artificial» y propuso relegar ese término al «montón de cenizas de la historia». «En realidad, lo que hoy llamamos inteligencia artificial no es artificial ni inteligente. Los primeros sistemas de IA estaban muy dominados por reglas y programa, de modo que, por lo menos, la palabra artificial estaba justificada. Pero los sistemas actuales, como el ChatGPT que tanto gusta a todos, no se basan en reglas abstractas, sino en el trabajo de seres humanos reales: artistas, músicos, programadores y escritores, de cuya obra creativa y profesional se apropian estos sistemas con la excusa de querer salvar la civilización. En todo caso, es una ‘inteligencia no artificial’».⁵⁸

En 2023, NOWOTNY, H. cuestionó que realmente la inteligencia artificial sea «inteligencia» y la descarta como el equivalente a «inteligencia humana». Añadió que la definición de la IA estuvo equivocada desde el principio: «Los expertos están de acuerdo en que hubiera sido mejor darle otro nombre, pero ahora es demasiado tarde».⁵⁹ Recalca que las máquinas digitales no entienden, no pueden dar sentido o conferir significado. No pueden tolerar la ambigüedad ni captar la ironía. «Pero son increíblemente efectivas para llevar a cabo las tareas que les encomendamos», dijo. Por contra, NOWOTNY recalca que la cognición humana (y animal) está «encarnada» porque todo el organismo está involucrado en la percepción, pensamiento, sentimiento, actuación y estar vivo. Hay una coordinación de mente y cuerpo.⁶⁰

Ese mismo año, DANS, E. también rechazó la definición y el concepto de inteligencia artificial pues prefiere referirse a ella como «machine learning» (aprendizaje profundo basado en procesos estadísticos avanzados desarrollados mediante herramientas en ocasión muy complejas y con una gran potencia de cálculo). Asegura que quienes «utilizan sin problemas ni remordimiento»⁶¹ la expresión inteligencia artificial suele ser porque «revelan, como dice el escritor de ciencia ficción Arthur C. Clarke en su Tercera Ley, su incapacidad para

⁵⁸ MOROZOV, E. , «Ni es artificial, ni es inteligente», 2 de abril de 2023, El País. Suplemento Ideas. pág. 5.

⁵⁹ PARRA, S., «Helga Nowotny», Muy Interesante, 504, 2023, pág. 39.

⁶⁰ Ídem, PARRA, S. (2023), pág. 39.

⁶¹ DANS, E., Todo vuelve a cambiar. Cómo la Web3 revolucionará el mundo tal y como lo conocemos, Deusto. Barcelona, 2023, pág. 171.

diferenciar una tecnología suficientemente avanzada de la magia»⁶². Añade que una máquina que piensa o una máquina inteligente no existe, pues «lo que hacen es, a partir de una serie de datos, reproducir los patrones que generan un resultado concreto y en un formato determinado»⁶³. Solo pueden hacer aquello para lo que han sido programadas y son ajenas, por mucho que puedan parecerlo cuando simulan determinadas conversaciones, de adquirir ningún tipo de supuesta conciencia»⁶⁴.

Otros autores como MOROZOV, E. (2023) siguen esa misma tendencia a cuestionar el concepto de inteligencia artificial, quien señala que «ni es inteligente ni es artificial»⁶⁵.

SANGUINETTI, P. (2023) también se hace eco de esas críticas respecto al término inteligencia artificial, que califica de «relato» o «narrativa» e incluso este autor plantea una nueva narrativa donde la IA no existe. Dado que el test de Turing o el juego de simulación se basa en la capacidad de las máquinas para engañar a un jurado, por lo tanto se trataría más bien de una máquina de «inteligencia simulada» porque el juego de la imitación reemplaza al pensamiento.⁶⁶ Cita a algunos críticos como LARSON, E.J., quien señala que es mejor hablar de «simulación de tareas humanas» o como LEUFER, D., que en su proyecto *AI Myths*⁶⁷, habla de que el nombre de la IA presenta un gran «malentendido»⁶⁸ y sugiere nombres alternativos como «estadística computacional», «automatización cognitiva», «optimización aplicada», «sistemas automáticos de decisión» o «programa de ordenador».⁶⁹

En concreto, *AI Myths* señala que existe un mito sobre el hecho de que el significado del término IA sea claro, lo que niega. Subraya que «la vaguedad del término ‘inteligencia artificial’ puede llevar a confusiones y mistificaciones con graves consecuencias». «Con cada avance genuino en el campo de la ‘inteligencia artificial’, vemos un aumento paralelo en exageraciones, mitos, conceptos erróneos e inexactitudes. Estos malentendidos contribuyen a la opacidad de los sistemas de IA, haciéndolos mágicos, inescrutables e inaccesibles a los ojos del pú-

⁶² Ídem, DANS, E. (2023), págs. 171-172.

⁶³ Ídem, DANS, E. (2023), págs. 171-172.

⁶⁴ Ídem, DANS, E. (2023), págs. 171-172.

⁶⁵ MOROZOV, E., «Ni es artificial, ni es inteligente». 2 de abril de 2023, El País, Suplemento Ideas, pág. 5.

⁶⁶ SANGUINETTI, P., «Tecnohumanismo», LHG, Madrid, 2023, págs. 56-59

⁶⁷ <https://www.aimyths.org/>

⁶⁸ Ídem, SANGUINETTI, P. (2023), pág. 61.

⁶⁹ Ídem, SANGUINETTI, P. (2023), pág. 61.

blico», dice esta web⁷⁰. En un post examinan la historia del término, cómo se usa hoy en día y sugieren algunas alternativas «para llamar a todo IA». ⁷¹ Propone términos alternativos como «programa de computadora elegante», «herramienta de aprendizaje automático», «máquina mágica de predicción basada en datos», «aplicación científicamente dudosa» o «aplicación propietaria sobrevalorada».

Una opacidad de los algoritmos o caja negra (black box) que, en el caso de estas tecnologías emergentes, puede afectar a los derechos fundamentales. Sentencias como la C-806/24 del Tribunal de Justicia de la UE (TJUE) consideran abusiva hacia el consumidor cuando las compañías usan la IA para tomar decisiones automatizadas (liquidación de facturas, por ejemplo) y no le dan explicaciones al cliente hasta que se celebra juicio, lo que obliga al usuario a afrontar las costas⁷².

SANGUINETTI, P. también cita al Center on Privacy & Technology⁷³ de la facultad de Derecho de la Universidad de Georgetown, que dejó de usar los términos «inteligencia artificial» y «machine learning» para «mitigar el perjuicio de las tecnologías digitales en la vida de individuos y comunidades», al constatar que muchas personas no saben lo que significan esos términos o porque sobredimensionan sus posibilidades.⁷⁴ Según TUCKER, E, en 2021, el Instituto FrameWorks, en asociación con la Fundación MacArthur, publicó un resumen de investigación describiendo algunos de los conceptos erróneos más frecuentes sobre qué es la «IA» y cómo funciona. Lo que encontramos más sorprendente en el Centro de privacidad sobre sus hallazgos fue que la mayoría de estos conceptos erróneos parecían caer en una de dos categorías generales, que son dos caras de la misma moneda: (1) el público no sabe qué es «IA». y (2) el público asume que la IA es más

⁷⁰ <https://www.aimyths.org/>

⁷¹ <https://www.aimyths.org/the-term-ai-has-a-clear-meaning>

⁷² Tribunal de Justicia de la Unión Europea, «Resumen de la petición de decisión prejudicial — Asunto C-806/24», presentado el 25 de noviembre de 2024 por el Sofiyski rayonenski sad (Tribunal de Primera Instancia de Sofía, Bulgaria).

Antes de presentar la demanda, el profesional no explicó al consumidor de forma clara y comprensible cómo se tomaron estas decisiones automatizadas. En consecuencia, previamente al procedimiento, el profesional no ha aclarado la manera en que se tomaron las decisiones automatizadas, sino que lo ha hecho durante el procedimiento por él mismo iniciado. El órgano jurisdiccional remitente considera que esta forma de proceder es abusiva. El profesional se sirve de su posición dominante para no proporcionar información acerca de la TDA antes de un procedimiento judicial.

<https://curia.europa.eu/juris/showPdf.jsf?text=inteligencia%2Bartificial%2B&docid=294955&pageIndex=0&doclang=ES&mode=req&dir=&occ=first&part=1&id=1241726>

⁷³ <https://www.law.georgetown.edu/privacy-technology-center/>

⁷⁴ Ídem, SANGUINETTI, P. (2023), pág. 61.

inteligente que ellos. Por ejemplo, en conversaciones con investigadores de FrameWorks sobre algoritmos predictivos, los participantes se centraron en la necesidad de proporcionar «buenos datos» al algoritmo, pero parecían no reconocer el algoritmo en sí mismo como algo construido. Al mismo tiempo, FrameWorks descubrió que «la gente piensa en los algoritmos predictivos como adivinación, que es la capacidad de ver el futuro y saber qué va a pasar en función de la información recopilada en el presente»⁷⁵. Esta reflexión fue publicada en el 2022 por TUCKER, E., directora ejecutiva del Centro de Privacidad y Tecnología de Georgetown Law en el artículo «Artificio e Inteligencia», que ella escribió en el blog Tech Policy Press de la institución. Dicho centro se comprometió a ser lo más específico posible sobre qué es la tecnología en cuestión y cómo funciona, identificar cualquier obstáculo para nuestra propia comprensión de una tecnología que resulte de fallas en la transparencia corporativa o gubernamental, nombrar a las empresas responsables de la creación y difusión del producto tecnológico y atribuir intención o agenda a los actores humanos que construyen y usan la tecnología, nunca a la tecnología misma.⁷⁶

A partir de las propuestas de la UE sobre la IA (el Libro Blanco o la AI Act del 2021), se empezó a considerar esta como un «producto», concepto que va ganando aceptación a partir de la firma de los tres principios asumidos por siete empresas y la Casa Blanca, con Joe Biden, el 21 de julio de 2023.⁷⁷

Hay que tener en cuenta que algunos autores definen la IA como una «inteligencia no humana», ya que estas máquinas de análisis masivos de datos no siguen los mismos patrones lógicos y de razonamiento que los seres humanos. Técnicamente, no es una inteligencia humana, aunque haya sido programada por humanos. Es lo más cercano que se podrá ver de una inteligencia «alien». Una de las razones es que estas máquinas no pueden explicar los métodos seguidos para obtener sus resultados, más fiables que los de los propios humanos y,

⁷⁵ <https://techpolicy.press/artifice-and-intelligence/>

⁷⁶ <https://techpolicy.press/artifice-and-intelligence/>

⁷⁷ Los «productos IA» son mencionados en los principios asumidos por las siete empresas de IA y que fueron formulados por Joe Biden en la Casa Blanca el 21 de julio de 2023. Viene enunciado en el epígrafe «Ensuring Products are Safe Before Introducing Them to the Public». FACT SHEET: «Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI», 21 de julio de 2023. <https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/>

por lo tanto, son inexplicables y opacos, o de difícil o imposible comprensión para un humano. Se pone, como ejemplo, que una máquina confundió la foto de un coche con un gato por el brillo en un píxel del faro. Estamos hablando de que un científico puede extraer conclusiones a partir del análisis de cinco o seis variables mientras que, como señala LEE, K-F. estas máquinas manejan miles de variables para obtener un resultado mucho más preciso.

ROBISCO, S. llega a una conclusión similar a de muchos de estos autores y explica por qué ella tampoco da una definición exacta de la inteligencia artificial: «No lo hemos hecho porque en la actualidad no existe una única definición de inteligencia artificial»⁷⁸. Alerta de que el concepto de IA se puede convertir en un «concepto vacío, sin ningún de sentido, si no especificamos a lo que nos referimos»⁷⁹. Aclara que las razones de la falta de una definición unificada «las encontramos en que el concepto de sistema inteligente ha cambiado mucho a lo largo de la historia y que además existen diversas ramas de pensamiento que definen la inteligencia artificial de forma diferente: unos basándose en los términos de fidelidad en la forma de actuar de los seres humanos y otras basándose en el concepto de racionalidad»⁸⁰. Para esta autora, el sistema de IA o «agente inteligente» es «aquel software que realiza la mejor acción posible ante una situación que ocurre dentro de un entorno cambiante».⁸¹

Respecto a estas definiciones formuladas a lo largo de casi 70 años, se deduce que existe una dualidad de interpretaciones. Por un lado, los autores aportan una definición tecnológica que apenas varía (una máquina o agente que hace tareas propias del comportamiento humano) y, por otro lado, le añaden una vertiente social (máquinas que proyectan las decisiones tomadas por un determinado grupo humano). El Derecho ha de sopesar ambos campos de definición para comprender el contexto en que se desarrolla la IA. A ello se suma una realidad y es que la tecnología de la IA está en plena evolución y sus técnicas son asumidas y superadas rápidamente según acontecen los avances, lo que obliga a una continua revisión y actualización de su definición.

La exposición de estas distintas definiciones sobre la IA a lo largo tres tercios de siglo evidencian que el problema de concretar este con-

⁷⁸ ROBISCO, S., Historia de la Inteligencia Artificial, editorial Guadalmazán, Madrid, 2024, pág. 24

⁷⁹ ROBISCO, S. (2024), «Historia de la In...», pág. 24

⁸⁰ Ídem, ROBISCO, S. (2024), pág. 24

⁸¹ Ídem, ROBISCO, S. (2024), pág. 24

cepto sigue vivo y lo seguirá pues en un futuro, como ya ha ocurrido recientemente con la IA generativa, la IA puede ser identificada, por ejemplo, con ChatGPT, un concepto novedoso e inexistente hasta entonces. No solo la definición ha evolucionado sino que algunos niegan la propia existencia del concepto de estudio. No se trata de probar cuál es la mejor definición o la más correcta de la IA sino de dejar en evidencia que no existe un acuerdo sobre lo que realmente es este concepto que, además, podrá variar en los próximos años hasta hacerlo irreconocible, en función de cómo avance el desarrollo de esta tecnología emergente y novedosa. Jurídicamente, las leyes de la IA se asientan sobre una definición del objeto sin concretar en profundidad, quizás porque el término «inteligencia», en general, tampoco está lo bastante claro desde el punto de vista científico. Si en el pasado no se le pudo achacar una mala construcción jurídica debido a que el término tuvo una rápida evolución, ahora, al menos, las instituciones son conscientes de los efectos que tendrá esta circunstancia a nivel jurídico. Una de las soluciones ha sido buscar una definición alternativa o un concepto más fácil de definir.

III. SISTEMAS DE IA

Frente al concepto de IA surge otro denominado «sistemas de IA» que, como ya se ha comentado surge con la primera definición del concepto en el Seminario de Dartmouth, y que ha sido adoptado por organismos internacionales como la UNESCO o por instituciones políticas como la UE. Pero estas entidades repiten el error de definir a un sistema de IA como «un sistema», incluyendo el objeto definido en la propia definición, lo que no se puede decir que contribuya a la claridad, con los importantes efectos jurídicos que, como ha ocurrido con la IA, tiene esta falta de concreción. Como se verá, gran parte de la definición de Dartmouth aflora en la AI Act.

La UNESCO se centró en definir los sistemas de IA como «sistemas capaces de procesar datos e información de una manera que se asemeja a un comportamiento inteligente, y abarca generalmente aspectos de razonamiento, aprendizaje, percepción, predicción, planificación o control».⁸² La UNESCO indica que los sistemas de IA son tecnologías de procesamiento de la información que integran modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales. Los

⁸² UNESCO, Documento 41C/73, Anexo, 2021, pág. 18.

sistemas de IA están diseñados para funcionar con diferentes grados de autonomía, mediante la modelización y representación del conocimiento y la explotación de datos y el cálculo de correlaciones. Entre los métodos incluye el aprendizaje automático, incluido el aprendizaje profundo y el aprendizaje de refuerzo. Añade también el razonamiento automático, incluidas la planificación, la programación, la representación del conocimiento y el razonamiento, la búsqueda y la optimización.⁸³

El Grupo de Expertos Independientes de Alto Nivel sobre la IA (AI HLEG) de la UE describió un sistema de IA⁸⁴ a través de tres capacidades principales: percepción / razonamiento, toma de decisiones y actuación. Dichos expertos creen que es suficiente para presentar y comprender la mayoría de las técnicas y subdisciplinas de IA que se utilizan actualmente para construir sistemas de IA, ya que todas incluyen diversas capacidades de los sistemas. En términos generales, AI HLEG agrupa todas estas técnicas en dos principales grupos que se refieren a la capacidad de razonamiento y aprendizaje. La robótica es otra disciplina muy relevante.

El grupo AI HLEG elaboró un documento separado en el que profundiza en la definición de sistemas de IA utilizada a efectos de las Directrices. Dicho documento se ha publicado en paralelo con el título «Una definición de la inteligencia artificial: Principales capacidades y disciplinas científicas»⁸⁵. El informe distingue entre IA estrecha (o débil) y general (o fuerte). Un sistema de IA general está destinado a ser un sistema que pueda realizar la mayor parte de las actividades que los humanos pueden hacer. Los sistemas de IA débiles son, en cambio, aquellos que pueden realizar una o pocas tareas específicas.

⁸³ UNESCO, «Proyecto de recomendación sobre la ética de la inteligencia artificial». Conferencia General, 41st, 2021, Documento 41C/73, Anexo, pág. 18.

⁸⁴ Ver documento en: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341

⁸⁵ COMISIÓN EUROPEA. Grupo de expertos de alto nivel sobre la IA. «Una definición de Inteligencia Artificial: principales capacidades y disciplinas científicas». Título original en inglés: «A definition of Artificial Intelligence: main capabilities and scientific disciplines», Publicado el 8 de abril de 2019. Ver el sitio en:

https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341 . Este documento amplía la definición de inteligencia artificial (IA) tal como se define en la Comunicación de la Comisión sobre IA. Aclara ciertos aspectos de la IA como disciplina científica y como tecnología, con el objetivo de evitar malentendidos, lograr un conocimiento común compartido de la IA que pueda ser utilizado de manera fructífera también por los no expertos en IA, y proporcionar detalles útiles que pueden ser utilizado en la discusión sobre las directrices de ética de la IA y las recomendaciones de las políticas de IA.

El Libro Blanco de la IA de la UE profundiza más sobre lo que el grupo de expertos entiende por «sistemas de IA» porque aportan una definición mucho más clara. En el COM(2018) 237 final, p. 1 viene recogido un concepto que parece más concreto: «El término «inteligencia artificial» (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción —con cierto grado de autonomía— con el fin de alcanzar objetivos específicos. Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas)».

Según la definición del grupo de expertos de alto nivel, (pág. 8): «Los sistemas de inteligencia artificial (IA) son programas informáticos (y posiblemente también equipos informáticos) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno mediante la adquisición de datos, la interpretación de los datos estructurados o no estructurados, el razonamiento sobre el conocimiento o el tratamiento de la información, fruto de estos datos y la decisión de las mejores acciones que se llevarán a cabo para alcanzar el objetivo fijado». ⁸⁶

La clave es que a lo largo del Libro Blanco y de sus menciones a los expertos de alto nivel, la UE equipara los sistemas de IA con «programas informáticos» o «posiblemente equipos informáticos», que es una forma bastante concreta de definir la IA. El Libro Blanco vuelve a incidir en ello al señalar que un programa informático es un «producto».

III.A. Tipos de sistemas de IA: IA General (fuerte), Débil (estrecha) y Singleton (extrema)

OROZCO⁸⁷ distingue dos sistemas de IA. El primero se refiere a aquellos sistemas de IA que se basan por completo en software, desenvolviéndose en el entorno virtual (por ejemplo, asistentes de voz, buscadores, sistemas de reconocimiento facial y de voz). El segun-

⁸⁶ También recogido en una nota al pie en la página 20 del Libro Blanco de la IA.

⁸⁷ OROZCO GONZÁLEZ, M., «Aspectos éticos de la inteligencia artificial: especial mención a las IA black box». Universidad de Murcia. Ponencia en el Congreso Internacional Hispano-Luso e Iberoamericano «Sociedad Digital y Derecho Civil», 13-15 abril de 2021. <https://sociedad-digital.es/>

do, comprende aquellos sistemas de IA integrados en dispositivos de hardware (coches autónomos, drones o aplicaciones de Internet de las Cosas).

Los sistemas de IA actualmente implementados son ejemplos de IA estrecha. En los primeros días de la IA, los investigadores utilizaron una terminología (IA débil y fuerte). Según el anterior informe de los expertos de la UE, aún quedan muchos desafíos éticos, científicos y tecnológicos abiertos para construir las capacidades que serían necesarias para lograr la IA general, como el razonamiento de sentido común, la autoconciencia y la capacidad de la máquina para definir su propio propósito.

En 2024, el Proyecto Final de la AI Act (Ley de la IA de la UE) definió el sistema de IA como «un sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía y que puede mostrar adaptabilidad después del despliegue y que, para objetivos explícitos o implícitos, infiere, a partir de la información que recibe, cómo generar resultados como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales».⁸⁸ La IA de Propósito General sería el equivalente a la IA fuerte o general.

DE LA HOZ, K. define la IA débil como «un sistema orientado a resolver problemas concretos y delimitados, algo que las máquinas aprenden a hacer a través de patrones y tendencias repetitivas gracias a algoritmos programados por humanos»⁸⁹. Pone como ejemplo, a asistentes virtuales como Siri (Apple), Alexa (Amazon) o Assistant (Google), que responden órdenes compatibles con una búsqueda en Internet o en su base de datos. Por contra, la IA fuerte sería «un sistema complejo capaz de abordar decisiones de forma proactiva, deductiva y autoconsciente». Se refiere a máquinas de las películas de ciencia-ficción como la computadora Hal 9000 de 2001. Una odisea del espacio (1969), los replicantes de Blade Runner (1982), Skynet de Terminator (1984), los agentes virtuales de The Matrix (1999), Her (2014), Ex Machina (2015) o The Creator (2023).

⁸⁸ Artículo 3 de la AIA actualizado el 21 de mayo de 2024 según la versión y numeración de artículos aprobada por el Parlamento de la UE el 14 de mayo de 2024. <https://www.euaiact.com/article/3>

⁸⁹ DE LA HOZ, K. y COELHO, F., «IA para periodistas. Una herramienta por explotar», Prodigioso Volcán, Madrid, 2021, págs. 15-16.

MINSKY distingue varios tipos de IA⁹⁰. En su forma más extrema, conocida como IA fuerte, el objetivo es construir una máquina capaz de pensamiento, conciencia y emociones. Esta visión mantiene que los humanos no son más que ordenadores elaborados. Por otro lado, la meta de la IA débil es desarrollar teorías de la inteligencia humana y animal, y probarlas construyendo modelos operativos, en la forma de programas de ordenador o robots.

BRIGHTON, H. y SELINA, H. identifican un tercer tipo de IA que sería la IA-alien⁹¹, la cual busca construir máquinas que no estén basadas necesariamente en la inteligencia humana o animal. Estas máquinas pueden mostrar comportamiento inteligente pero la base para este comportamiento no es importante. Los mecanismos que subyacen en estos sistemas no pretenden copiar a la inteligencia humana.

MERABET, S.⁹² distingue entre una inteligencia disminuida y una inteligencia aumentada. En el primer caso, la IA carece de elementos humanos como la ausencia de consciencia, lo que excluye, a nivel jurídico, la preminencia de la voluntad y la falta de sentimientos. En este tipo de inteligencia artificial, el autor cree que es incompatible con los principios directores de un proceso judicial o justicia predictiva.

Respecto a la segunda versión de la IA, la inteligencia aumentada, MERABET, S. teme que los excesos atenten contra las personas, por ejemplo, por la invasión de la vida privada o al libre arbitrio, así como la aplicación de la norma.

Por su parte, TEGMARK, M.⁹³ diferencia entre distintos tipos de inteligencia⁹⁴: la inteligencia (capacidad de alcanzar objetivos complejos), la inteligencia artificial (inteligencia no biológica), la inteligencia estrecha o débil (capacidad de alcanzar un conjunto limitado de objetivos; por ejemplo, jugar al ajedrez o conducir un coche), la inteligencia general (capacidad de alcanzar prácticamente cualquier objetivo, incluido el aprendizaje), la inteligencia universal (capacidad de adquirir inteligencia general a partir del acceso de datos y recursos), inteligencia artificial general (capacidad para realizar cualquier tarea cognitiva al menos tan bien como los humanos), IA de nivel humano (IAG), IA fuerte (IAG), Superinteligencia (ASI, inteligencia

⁹⁰ Citado por BRIGHTON, H. y SELINA, H., *Inteligencia Artificial. Una guía ilustrada*, Tecnos, Madrid, 2019, págs. 14-15

⁹¹ BRIGHTON, H. y SELINA, H. (2019), pág. 16.

⁹² MERABET, S. (2020), págs. 193-258.

⁹³ TEGMARK, M. (2018), págs. 56-57.

⁹⁴ TEGMARK, M., *Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial*, Penguin Random House Grupo Editorial, Barcelona, 2018, págs. 56-57.

general de nivel muy superior al humano) e IA amigable (Superinteligencia cuyos objetivos coinciden con los nuestros). Una última categoría sería la IA segura (una IA cuyos problemas de seguridad han sido resueltos)⁹⁵.

Según el NSCAI⁹⁶, algunos investigadores usan la frase «inteligencia artificial general» (AGI o IAG) para referirse al objetivo de extender la IA más allá de las cuñas verticales y estrechas de la experiencia. Los debates se han centrado en si fuese posible que concurriesen avances específicos que conduzcan a capacidades más generales, similares a las humanas, o si es más probable que el campo continúe impulsando una IA más general a lo largo de una o más dimensiones de habilidades. Independientemente de la perspectiva, ven que el progreso es significativo.

SCHNEIDER, S. señala que la IAG es una inteligencia que, como la humana, puede combinar conocimientos de diferentes áreas temáticas y demostrar flexibilidad y sentido común. Advierte que está previsto que la IAG deje obsoletas muchas profesiones humanas en las próximas décadas. Según MULLER y BOSTROM (2016) la IA ejercerá la mayor parte de las profesiones humanas al menos tan bien como un humano medio, con un 50 % de probabilidades antes de 2050 y con un 90% de probabilidades antes de 2070.⁹⁷

Otros autores como SHANAHAN, M.⁹⁸ distinguen la IA por su grado de consciencia. Las actuales inteligencias artificiales y robots puntúan cero respecto a las cualidades humanas y la capacidad de consciencia, pero algunas IA que tienen capacidad de movimiento y adaptación logran una puntuación mejor. En cambio, un ente creado como sucesor de los humanos (que denomina «hijos mentales») tendría un grado de consciencia alta y cualidades humanas significativas. Añade que los seres exóticos conscientes, al igual que las inteligencias extraterrestres, quizás sean más conscientes que ellos, pero también distintos.

⁹⁵ TEGMARK, M., Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial, Penguin Random House Grupo Editorial, Barcelona, 2018, pág. 59.

⁹⁶ NSCAI, Informe, pág. 30.

⁹⁷ Referencia: MÜLLER, V.C. Y BOSTROM, N., «Future Progress in Artificial Intelligence: A Survey of Expert Opinion», en MÜLLER, V.C., en *Fundamental Issues of Artificial Intelligence*, Synthese Library, Springer, Berlín, 2014.

⁹⁸ Citado por HENDERSON, C., *El mapa de las maravillas*, Ático de los libros, Barcelona, 2021, pág. 385

Según CASILLI, A.A., el sector denomina en su jerga, de forma irónica, «inteligencia artificial artificial»⁹⁹ a los miles de autoempleados de plataformas de micro tareas como Amazon Mechanical Turk que clasifican a mano las direcciones o categorizan en función de su pertinencia todos los productos del catálogo. Este trabajo previo, explica CASILLI, como proporcionar ejemplos para entrenar al software, y posterior, como verificar que sus resultados son correctos, es esencial para el éxito comercial de Amazon y las compañías de su ecosistema. Cada actividad de un «turker» se denomina «tarea de inteligencia humana» (Human Intelligence Task, o HIT)¹⁰⁰.

Dentro de este contexto, CASILLI encuadra esta actividad manual dentro de la inteligencia artificial débil, que incorpora un importante componente de trabajo de inteligencia no artificial.¹⁰¹ Para entender la presencia de trabajadores humanos en tareas automatizadas que corresponden a soluciones de IA orientadas al cliente hay que tener en cuenta que estas se han externalizado y el trabajo se ha parcelado (taylorización). No es una inteligencia fuerte que intente sustituir la acción humana por un «supercerebro» sino que se trata de usuarios que usan dispositivos con IA débiles, creados a partir de aplicaciones que ayudan a gestionar la información (retoques de imagen, clasificaciones, determinar la mejor ruta, o la ayuda a los deberes de los niños que presta el software de bots conversaciones como Siri o Alexa). Este tipo de IA débil incorpora a un fuerte componente de trabajo no artificial.¹⁰² Parece que el llamado «digital labor» no es una cuestión baladí y que habrá que tener en cuenta a la hora de dirimir responsabilidades o de evidenciar que no es posible dirimirla ante los miles de microtrabajadores que pueden estar implicados en una aplicación de IA (agentes virtuales asistidos por humanos, HAVA¹⁰³) o dispositivos «impulsados por humanos».

La conclusión de CASILLI es que el «fantasma de la IA fuerte» está cediendo paso a la IA débil y supervisada por humanos, la cual considera que es la «única inteligencia posible», limitada e ineficaz en ausencia de la intervención humana. Más allá de los ingenieros que configuran la IA fuerte, lo que le llama la atención son «los miles de millones de manos pequeñas que, día a día, operan la marioneta de la

⁹⁹ CASILLI, A.A., *Esperando a los robots. Investigación sobre el trabajo del clic*, Punto de Vista Editores, Madrid, 2021, págs. 126-128.

¹⁰⁰ Ídem, CASILLI (2021), pág. 128.

¹⁰¹ Ídem, CASILLI (2021), págs. 56-57.

¹⁰² Ídem, CASILLI (2021), pág. 57.

¹⁰³ HAVA: Human-Assisted Virtual Agents.

automatización débil»¹⁰⁴. Si las IA no están completamente automatizadas, «surge la duda de que no lo estén en absoluto».

LATORRE, J.I. menciona cierto tipo de inteligencia artificial avanzada única (que en inglés se denomina Singleton), un tipo de IA extrema vastamente superior a la humana que sería suprema y con características benévolas: pacífica y nada hostil porque no hallaría beneficio en la violencia. Si surgiese otra inteligencia artificial, ambas optarían por unirse y generar una inteligencia más potente,¹⁰⁵ y no le cabe duda de que una inteligencia superior será pacífica (porque procede de una sociedad educada) y el comportamiento violento sería ajeno a su código fuente. Además, serían máquinas tolerantes y con respeto absoluto por la diversidad de los comportamientos humanos (inferir el mejor comportamiento para hacer feliz al máximo de humanos, con el peligro de que la máquina no tolere las disidencias y la generación de dilemas ante el mal comportamiento humano con el medio ambiente), levemente inconsistentes y buscadora del conocimiento absoluto. LATORRE avisa de que es fácil programar en una máquina que finja tener una duda al escoger una expresión (lo que la hace más humana) y codificar el amor, el odio, la indiferencia o la pasión. Y cree que estas inteligencias artificiales supremas han de ser libres.¹⁰⁶

Este debate sobre la Inteligencia Artificial General (AGI) y el peligro de que asuma el control de la Humanidad es descartado por el informe «Preparación para el futuro de la inteligencia artificial»¹⁰⁷ elaborado por la Administración Obama. Según COECKELBERGH, M. el informe afirma que las preocupaciones a largo plazo sobre una IAG superinteligente «deberían tener poco impacto en la política actual» porque ve más preocupaciones actuales del aprendizaje automático como el sesgo o el hecho de que los desarrolladores no comprendan su sistema lo suficientemente bien para evitar determinadas consecuencias.¹⁰⁸

¹⁰⁴ Ídem, CASILLI (2021), pág. 59.

¹⁰⁵ LATORRE, J.I., *Ética para máquinas*, Ariel, Barcelona, 2019, págs. 242-243.

¹⁰⁶ Ídem, LATORRE, J.I. (2019), págs. 278-291.

¹⁰⁷ OBAMA, B. Executive Office of the President, National Science and Technology Council Committee on Technology, «Preparing for the Future of Artificial Intelligence», octubre de 2016.

https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

¹⁰⁸ COECKELBERGH, M., *Ética de la Inteligencia Artificial*, Cátedra, Madrid, 2021, pág. 127.

No obstante, los borradores del Código de Buenas Prácticas de la IA de Propósito General han debatido el riesgo sistémico de la IA (contemplado en la AI Act) con el fin de recomendar al fabricante que realice ensayos y pruebas antes de comercializar productos que generen problemas de forma masiva.

IV. LAS LAGUNAS EN LA AI ACT

El artículo tercero de la AI Act, aprobada en agosto del 2024, define el «sistema de IA», a efectos del reglamento, como «un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales»¹⁰⁹. Como se puede comprobar, parece una versión actualizada del concepto original del Seminario de Dartmouth, lo que lleva a preguntarse para qué han servido tantas décadas de debate entre autores sobre el concepto de IA. La definición tampoco recoge las pruebas aportadas por CRAWFORD, K. sobre el complejo ecosistema extractivo que subyace tras la IA, que más que un sistema es un conglomerado que, probablemente, la UE tendría que regular más profundamente.

Como se puede observar, el reglamento no define la IA sino los sistemas de IA, al igual que hicieron en el seminario de Dartmouth, y que se considera el «pecado capital» de la IA, que es equiparar el funcionamiento de una máquina con una cualidad humana como es la inteligencia. En su considerando primero, la AI Act menciona la IA por primera vez como «la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable» y luego garantiza «la libre circulación transfronteriza de mercancías y servicios basados en la IA» pero sin concretar a qué se refiere exactamente. Añade en el considerando sexto que «la IA debe ser una tecnología centrada en el ser humano. Además, debe ser una herramienta para las personas y tener por objetivo último aumentar el bienestar humano». Para entender a

¹⁰⁹ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial)

https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=OJ:L_202401689#enc_1

qué se refiere con IA hay que remitirse al Libro Blanco y al comité de expertos de alto nivel, que equiparaban la «IA» con un «producto», en concreto «programas informáticos».

La explicación de por qué se eligió esta definición de sistema de IA está recogida en el considerando 12.¹¹⁰ Dicho considerando indica que debe definirse con claridad el concepto de «sistema de IA» en el Reglamento y armonizarlo estrechamente con los trabajos de las organizaciones internacionales que se ocupan de la IA, a fin de garantizar la seguridad jurídica y facilitar la convergencia a escala internacional y una amplia aceptación, al mismo tiempo que se prevé la flexibilidad necesaria para dar cabida a los rápidos avances tecnológicos en este ámbito.

Además, señala el mismo considerando, la definición debe basarse en las principales características de los sistemas de IA que los distinguen de los sistemas de software o los planteamientos de programación tradicionales y más sencillos, y no debe incluir los sistemas basados en las normas definidas únicamente por personas físicas para ejecutar automáticamente operaciones.

El considerando 12 recalca que una característica principal de los sistemas de IA es su capacidad de inferencia, la cual se refiere al proceso de obtención de resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos y virtuales, y a la capacidad de los sistemas de IA para deducir modelos o algoritmos, o ambos, a partir de información de entrada o datos. Las técnicas que permiten la inferencia al construir un sistema de IA incluyen estrategias de aprendizaje automático que aprenden de los datos cómo alcanzar determinados objetivos y estrategias basadas en la lógica y el conocimiento que infieren a partir de conocimientos codificados o de una representación simbólica de la tarea que debe resolverse. La capacidad de inferencia de un sistema de IA trasciende el tratamiento básico de datos, al permitir el aprendizaje, el razonamiento o la modelización.

El término «basado en una máquina», señala el considerando, se refiere al hecho de que los sistemas de IA se ejecutan en máquinas.

La referencia a objetivos explícitos o implícitos subraya que los sistemas de IA pueden funcionar con arreglo a objetivos definidos ex-

¹¹⁰ Considerando 12 de la «Ley de Inteligencia Artificial (Reglamento (UE) 2024/1689), versión del Diario Oficial de 13 de junio de 2024». Expediente interinstitucional: 2021/0106(COD).

<https://artificialintelligenceact.eu/es/recital/12/>

plícitos o a objetivos implícitos. Los objetivos del sistema de IA pueden ser diferentes de la finalidad prevista del sistema en un contexto específico.

A los efectos del Reglamento, prosigue el Considerando 12, debe entenderse por entornos los contextos en los que funcionan los sistemas de IA, mientras que los resultados de salida generados por el sistema de IA reflejan las distintas funciones desempeñadas por los sistemas de IA e incluyen predicciones, contenidos, recomendaciones o decisiones.

Los sistemas de IA están diseñados para funcionar con distintos niveles de autonomía, lo que significa que pueden actuar con cierto grado de independencia con respecto a la actuación humana y tienen ciertas capacidades para funcionar sin intervención humana. La capacidad de adaptación que un sistema de IA podría mostrar tras su despliegue se refiere a las capacidades de autoaprendizaje que permiten al sistema cambiar mientras está en uso.

Los sistemas de IA, concluye el considerando 12, pueden utilizarse de manera independiente o como componentes de un producto, con independencia de si el sistema forma parte físicamente del producto (integrado) o contribuye a la funcionalidad del producto sin formar parte de él (no integrado).

Si bien el considerando 12 aclara los conceptos que compondrán el artículo 3.1 de la AI Act no explica exactamente qué es un «sistema», si se refiere a un programa o a una red de programas o dispositivos, aunque bastará con remitirse al Libro Blanco para entender que se refiere a «programas informáticos», en línea con lo expuesto por el grupo de expertos de alto nivel. Esta concreción previa en el Libro Blanco tiene relevancia, por ejemplo, cuando un sistema de IA (compuesto de numerosos nodos y dispositivos) coordina numerosas operaciones en tiempo real y que deben ser ejecutadas en red de forma coordinada (el sinónimo sería el concepto médico de sistema respiratorio o circulatorio). El considerando no aclara la naturaleza concreta del «sistema de IA» sino que se remite a que está basado en máquinas. En otras palabras, de no tener la referencia del Libro Blanco de la IA, tendríamos que imaginar a qué se refiere exactamente con «sistema de IA» y a pesar de que el propio considerando anima al Reglamento a definirlo «con claridad».

Otra cuestión que es interesante es que este considerando 12 aboga por algún tipo de unificación de la definición que sea común y que la puedan operar distintos organismos internacionales. En otras palabras, la definición de IA o de «sistemas de IA», de no ser por el

Libro Blanco, estaría sin concretar pero, contradictoriamente, el mismo texto aboga por mejorarlo (o buscar una definición consensuada a nivel internacional), por lo que mantiene vivo el problema de la inseguridad jurídica, algo que, debido a la constante innovación de esta disciplina, podría convertirse en un problema crónico a nivel jurídico en lo que respecta a esta disciplina.

A partir de dicho considerando, el Reglamento define el «sistema de IA», en el artículo 3.1. de la AI Act¹¹¹ y sobre el cual va a girar todo el documento jurídico y cuya definición ha sido expuesta al inicio de este epígrafe. La AI Act cita características como la autonomía (una de las características enunciadas por Dartmouth), la capacidad de adaptación y la inferencia (predicciones, contenidos, recomendaciones o decisiones, con lo que engloba todos los efectos conocidos de las redes sociales y de la IA generativa).

Sin embargo, la eficacia normativa de artículo 3.1 de la AI Act adquiere su mayor valor, desde el punto de vista jurídico, al considerar que estos sistemas de IA pueden «influir en entornos físicos o virtuales». Ello la convierte en una tecnología que también es autónoma y genera responsabilidades por daños y perjuicios desde el momento en que interactúa con su entorno.

Aunque los autores examinados anteriormente diferenciaban entre IA débil o IA fuerte o IA General (AGI), el Reglamento ha optado por centrarse en la IA generativa, considerada como un modelo de uso general de la IA, que en el capítulo 3.63 lo define como «un modelo de IA, también uno entrenado con un gran volumen de datos utilizando autosupervisión a gran escala, que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas, independientemente de la manera en que el modelo se introduzca en el mercado, y que puede integrarse en diversos sistemas o aplicaciones posteriores, excepto los modelos de IA que se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su introducción en el mercado».

Y a su vez, el capítulo 3.66 define el sistema de IA de uso general como «un sistema de IA basado en un modelo de IA de uso general y que puede servir para diversos fines, tanto para su uso directo como para su integración en otros sistemas de IA».

La solución que proponemos para resolver esta supuesta laguna legal de la AI Act es remitirnos a las Directrices éticas para una IA fia-

¹¹¹ <https://artificialintelligenceact.eu/es/article/3/>

ble¹¹², documento elaborado por el Grupo Independiente de Expertos de Alto Nivel de Inteligencia Artificial creado por la Comisión Europea en 2018. El punto 144 define la IA como «una disciplina científica que incluye varios enfoques y técnicas, como el aprendizaje automático (del que el aprendizaje profundo y el aprendizaje por refuerzo constituyen algunos ejemplos), el razonamiento automático (que incluye la planificación, programación, representación y razonamiento de conocimientos, búsqueda y optimización) y la robótica (que incluye el control, la percepción, sensores y accionadores así como la integración de todas las demás técnicas en sistemas ciberfísicos)».

En el punto 145, el Grupo de Expertos aclara que ha elaborado un documento separado en el que profundiza en la definición de sistemas de IA y que se titula «Una definición de la inteligencia artificial: Principales capacidades y disciplinas científicas». Dicho documento indica en su primer párrafo: «Partimos de la siguiente definición de inteligencia artificial (IA) propuesta en la Comunicación de la Comisión Europea sobre la IA: El término «inteligencia artificial» (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción —con cierto grado de autonomía— con el fin de alcanzar objetivos específicos. Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas).»¹¹³

El grupo de expertos continúa diciendo: «El término IA contiene una referencia explícita al concepto de inteligencia. Ahora bien, dado que la inteligencia (sea en las máquinas o en los seres humanos) es un concepto vago pese a haber sido estudiado en profundidad por psicólogos, biólogos y neurocientíficos, los investigadores que trabajan en el terreno de la IA utilizan principalmente el término «racionalidad». La racionalidad hace referencia a la capacidad de elegir la mejor acción posible para alcanzar un objetivo determinado, dados determinados criterios que es necesario optimizar y teniendo en cuenta los

¹¹² Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías y Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Directrices éticas para una IA fiable, Oficina de Publicaciones, 2019, <https://data.europa.eu/doi/10.2759/14078>

¹¹³ «Una definición de la inteligencia artificial: Principales capacidades y disciplinas científicas»

<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

recursos disponibles. Por supuesto, la racionalidad no es el único ingrediente del concepto de inteligencia, pero sí uno muy importante. En este documento utilizaremos la expresión sistema de IA para referirnos a cualquier componente, software o hardware basado en IA. De hecho, los sistemas de IA suelen estar integrados en forma de componentes de sistemas más amplios; no es tan frecuente encontrarlos como sistemas autónomos. Por lo tanto, un sistema de IA es, ante todo, racional, de acuerdo con uno de los manuales de inteligencia artificial más utilizados».

Dado que esta definición se elaboró en torno al año 2018, parte de la definición de la IA con la que se va a regular la UE podría haber quedado obsoleta. En aquel momento, se mencionaba a la IA como un sistema de aprendizaje profundo pero, cinco años después, la IA generativa es mucho más potente y más popular, ya que muchos usuarios la llevan instalada en su teléfono móvil como una aplicación de ChatGPT o de DeepSeek. La definición no tiene en cuenta tampoco los chatbots de IA emocional, que pueden manipular al usuario.

La cuestión es que, aunque se incluyesen las recientes innovaciones en una definición actualizada de la IA, la UE necesitaría hacer nuevas revisiones periódicas para añadir las novedades, por ejemplo por la IA combinada con ordenadores cuánticos, y, de esta forma, evitar posibles lagunas legales.

Es un problema que, como vemos, se remonta a 1956 y que la UE ha heredado al incorporar los principales características de la primera definición de Dartmouth (autonomía, adaptación al entorno, búsqueda de objetivos), actualizando el concepto con las innovaciones tecnológicas e ignorando las críticas de los autores. Si bien la definición de la IA casi 70 después incluye importantes novedades como el aprendizaje profundo que en los años 50 del siglo XX no existía también es cierto que, previsiblemente, en una década la IA actual habrá sido superada ampliamente por avanzados modos de razonamiento digital que ahora desconocemos y que obligará a una revisión del concepto de IA o sistema de IA tal y como fue planteada hace siete décadas.

V. CONCLUSIONES

La IA, como se ha probado en el primer capítulo referido a la definición de esta tecnología, evoluciona constantemente, de forma que una regulación sobre esta disciplina puede quedarse obsoleta fácilmente en cuestión de meses o años. Prueba de ello es el surgimiento

de ChatGPT, una IA generativa que habría dejado obsoleta la AI Act de no haberse incluido la categoría de IA de Propósito General antes de su aprobación. Podría pasar lo mismo con la IA multimodal, aplicada a los teléfonos móviles o la smart TV, o con cualquier otra tecnología similar. Esta continua evolución, que es imprevisible dado el continuo avance de la «lex artis», tendría que convertirse en una característica consustancial de la legislación de la IA, de modo que una comisión actualizase periódicamente las novedades y dictase la correspondiente regulación, que sería añadida al «corpus» legal. En parte, es lo que se está intentando a través del Código de Buenas Prácticas de la IA de Propósito General que, en el 2024 y 2025, se centró en elaborar instrucciones para los fabricantes que lancen al mercado productos de IA que deban respetar los derechos de autor y propiedad intelectual o que supongan un riesgo sistémico o de ciberseguridad. Esta evaluación continua está para quedarse, al menos hasta que todo el potencial de desarrollo de la IA quede agotado, lo que no parece inminente sino que puede prolongarse durante décadas. Esta inestabilidad jurídica por una falta de definición actualizada de la IA podría afectar negativamente a los derechos fundamentales de los ciudadanos cada vez que surja una innovación que genere una laguna legal y el legislador no actúe con agilidad o no comprenda cómo funciona dicha novedad técnica. Y hay ejemplos recientes como el tiempo que llevó en comprender el funcionamiento de las redes sociales y la economía de la captación de la atención del usuario y entender por qué el negocio estaba en extraer los datos personales del usuario. Cuando el legislador lo entendió, redactó el Reglamento General de Protección de Datos (RGPD), texto vinculado a la AI Act y al que remite en su introducción.

Alcanzar un acuerdo para una definición única de la IA supondría establecer una regulación global de la IA, en parte como pretenden la UNESCO, con sus Recomendaciones, o la ONU, y a pesar de que una división entre bloques (Estados Unidos y China) genera dos modelos distintos de IA. Esto obliga como mínimo a una serie de acuerdos puntuales en materia de seguridad o de derechos fundamentales, de obligado cumplimiento en toda la industria, y que en parte han sido recogidos a través del concepto de IA fiable y segura.

Una propuesta de este artículo es la creación y mantenimiento de una comisión de expertos o representantes políticos que actualice constantemente la legislación, como se ha hecho con el Código de Buenas Prácticas de la IA del 2025, para actualizar el Reglamento y adaptarlo a los continuos avances técnicos en esta tecnología emergente. Una idea sería que la Comisión Europea añadiese nuevos capítulos al texto

principal de la regulación, a medida que surjan novedades técnicas y se aprecie o se intuya su problemática, para crear un corpus legal único y actualizado, a fin de mitigar la inseguridad jurídica.

Todo apunta a que la definición de IA tendrá que seguir siendo actualizada periódicamente para evitar que la legislación se quede desfasada y facilite la creación de lagunas legales. O, probablemente, haya que replantearse la definición heredada del Seminario de Dartmouth para ampliarla e introducir un contexto social que ayude a comprender la complejidad del ecosistema de la IA. Para mayor seguridad jurídica, sería necesaria una comprensión más profunda de la IA y lo que significa así como una actualización de la definición que enfoque mejor el problema social que conlleva para tener una herramienta más precisa con la que salvaguardar los derechos fundamentales del ciudadano y a los que la IA, en algunos casos, ha afectado a derechos como la privacidad o la no discriminación, efectos negativos que la legislación está intentando mitigar.

VI. BIBLIOGRAFÍA

- AGRAWAL, A., Gans, J. y Goldfarb, A. Máquinas predictivas. La economía simple de la Inteligencia Artificial, Editorial Reverte, Barcelona, 2019.
- AYLLÓN GARCÍA, J.A., «La inteligencia artificial como un medio para administrar Justicia», en Fodertis 8.0. Estudios sobre tecnologías disruptivas y justicia, Editorial Comares, Granada, 2020.
- BODEN, M.A., Inteligencia artificial, Turner Publicaciones, Madrid, 2022.
- BRIGHTON, H. y SELINA, H., Inteligencia Artificial. Una guía ilustrada, Tecnos, Madrid, 2019.
- CABALLERO, R. y MARTÍN, E., Las bases de big data y de la inteligencia artificial, Los libros de la Catarata, Madrid, 2022.
- CASILLI, A.A., Esperando a los robots. Investigación sobre el trabajo del clic, Punto de Vista Editores, Madrid, 2021.
- CASTETS-RENARD, C., Droit du marché unique numérique et intelligence artificielle. «Droit de l'Union européenne. Manuels». Editorial Bruylant, Bruselas, 2020.
- COECKELBERGH, M., Ética de la Inteligencia Artificial, Cátedra, Madrid, 2021.

- COMISIÓN EUROPEA. Grupo de expertos de alto nivel sobre la IA. «Una definición de Inteligencia Artificial: principales capacidades y disciplinas científicas». Título original en inglés: «A definition of Artificial Intelligence: main capabilities and scientific disciplines», Publicado el 8 de abril de 2019.
- CRAWFORD, K., Atlas of AI, Yale University Press, New Haven y Londres, 2021.
- DANS, E., Todo vuelve a cambiar. Cómo la Web3 revolucionará el mundo tal y como lo conocemos. Barcelona. Deusto, 2023.
- DE LA HOZ, K. y COELHO, F. (2021), «IA para periodistas. Una herramienta por explotar», Prodigioso Volcán, Madrid,
- ERTEL, W., Introduction to Artificial Intelligence, Springer, Cham (Suiza), 2011, 2017.
- FERNÁNDEZ CARBALLO-CALERO, La propiedad intelectual de las obras creadas por inteligencia artificial, Aranzadi, Cizur Menor, Navarra, 2021.
- HENDERSON, C., El mapa de las maravillas, Ático de los libros, Barcelona, 2021.
- KAPLAN, J., Inteligencia Artificial. Lo que todo el mundo debe saber, Teell Editorial, 2017.
- LATORRE, J.I., Ética para máquinas, Ariel, Barcelona, 2019.
- MENECEUR, Y., L'intelligence artificielle en procès, Editorial Bryulant, Bruselas, 2020.
- MERABET, S., Vers un droit de l'intelligence artificielle. Nouvelle Bibliothèque de Thèses, v. 197, Éditions Dalloz, París, 2020.
- MOROZOV, E., «Ni es artificial, ni es inteligente». 2 de abril de 2023. El País. Suplemento Ideas.
- MÜLLER, V.C. Y BOSTROM, N., «Future Progress in Artificial Intelligence: A Survey of Expert Opinion», en MÜLLER, V.C., en Fundamental Issues of Artificial Intelligence, Synthese Library, Springer, Berlín,
- NIÑO-BECERRA, S., Capitalismo (1679-2065), Editorial Ariel, Planeta, Barcelona, 2020.
- OROZCO GONZÁLEZ, M., «Aspectos éticos de la inteligencia artificial: especial mención a las IA black box». Universidad de Murcia. Ponencia en el Congreso Internacional Hispano-Luso e Iberoamericano «Sociedad Digital y Derecho Civil», 13-15 abril de 2021.

- PANIAGUA, E., Error 404. ¿Preparados para un mundo sin Internet?, Penguin Random House Grupo Editorial, Barcelona, 2021.
- ROBISCO, S., Historia de la Inteligencia Artificial, Guadalmezán, Madrid, 2024
- ROSSI, F., «AI Ethics for Enterprise AI». IBM Research AI, 2019.
- RUSSELL, S., Human Compatible. AI and Problem of Control, Penguin, Gran Bretaña, 2020.
- RUSSELL, S. Y NORVIG, P., Artificial Intelligence: A Modern Approach, Prentice Hall, New Jersey. Tercera edición, 2009.
- SANGUINETTI, P., «Tecnohumanismo», LHG, Madrid, 2023.
- TEGMARK, M., Vida 3.0. Qué significa ser humano en la era de la inteligencia artificial, Penguin Random House Grupo Editorial, Barcelona, 2018.
- TURING, A.M., «Computing Machinery and Intelligence», Mind, Volume LIX, Issue 236, 1 de octubre de 1950, págs. 433–460.
- UE. «Libro blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza». Bruselas, 19.2.2020 / COM(2020) 65. [Primer párrafo]. Enlace al texto en: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf
- UE. Una Europa Adaptada a la Era Digital: inteligencia artificial (Ley de IA / AI Act). https://ec.europa.eu/commission/presscorner/api/files/document/print/es/ip_21_1682/IP_21_1682_ES.pdf
- UNESCO, «Proyecto de recomendación sobre la ética de la inteligencia artificial». Conferencia General, 41st, 2021, Documento 41C/73, Anexo.
- WEBB, A., Los nueve gigantes. Cómo las grandes tecnológicas amenazan el futuro de la humanidad, Ediciones Península, Barcelona, 2021.

